

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **05224694 A**

(43) Date of publication of application: 03 . 09 . 93

(51) Int. Cl. **G10L 3/02**(21) Application number: **04027906**(71) Applicant: **RICOH CO LTD**

(22) Date of filing: 14 . 02 . 92

(72) Inventor: **KURIKI SHOJI**(54) **SPEECH RECOGNITION DEVICE**

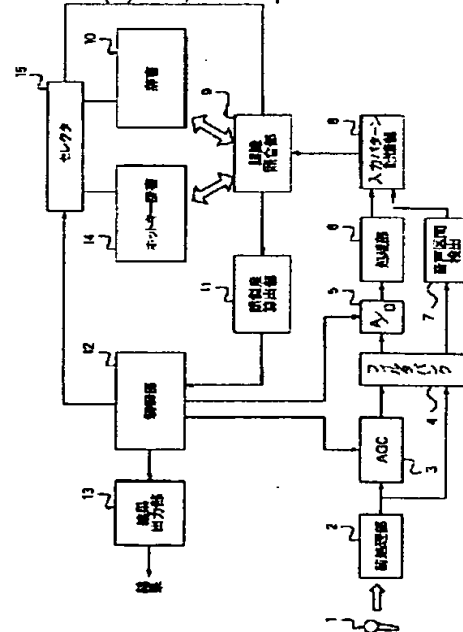
recognition rate.

(57) Abstract:

COPYRIGHT: (C)1993,JPO&Japio

PURPOSE: To obtain the largest recognition rate regardless of whether the speech level of a speaker is high or low by judging the speech level of an inputted hot key and setting the gain of the whole system corresponding to the judged level.

CONSTITUTION: An speech signal which is inputted is amplified by a microphone amplifier, divided into respective frequency bands by a filter bank 4, and digitized, and then, a process part 6 extracts features. The feature-extracted speech pattern is divided into speech sections, which are stored in an input pattern storage part 8. When the stored speech signal is inputted to a recognition and collation part 9, it is judged that a hot key process is completed; when the hot key is inputted, a selector 15 selects one of hot key dictionaries 14 provided by gains and the recognition and collation part 9 performs collation by using this dictionary to detect the dictionary which gives maximum similarity. The speech level of the speaker is estimated from the gain of this dictionary and the speech level of the whole system is so set as to obtain the maximum



(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)特許公開番号

特開平5-224694

(43)公開日 平成5年(1993)9月3日

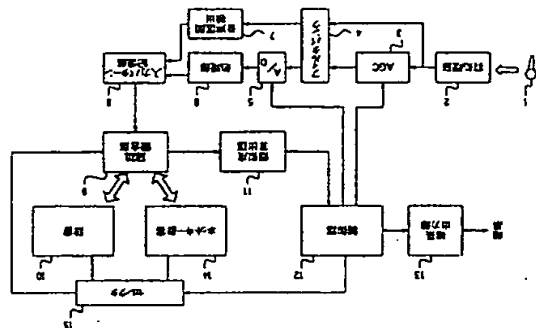
(51)Int.Cl. ⁴	G10L 3/02	識別記号	301 B 8842-5H	庁内整理番号	FI	技術表示箇所
(21)出願番号	特開平4-27806	(71)出願人	00006747	審査請求 未請求	請求項の数4(全9頁)	
(22)出願日	平成4年(1992)2月14日	株式会社リコー	東京都大田区中馬込1丁目3番6号	(72)発明者	栗本 章次	東京都大田区中馬込1丁目3番6号 株式会社リコー内
(74)代理人	井理士 齊山 保 (外1名)					

(54)【発明の名称】 音声認識装置

(57)【要約】

【目的】 発声者の音声レベルの高値にかかわらず、最大振幅が得られる音声認識装置を提供すること。

【構成】 ホットキー入力を受けて動作を開始するようにした音声認識装置において、入力されたホットキーの音声レベルを判定し、判定された音声レベルに応じたシステム全体のゲインを設定し、常に最大振幅が得られるようにする。



【特許請求の範囲】

【請求項1】 マイクロホンと、マイクrohンの音声信号出力を増幅するマイクアンプと、マイクアンプの出力信号をデジタル信号に変換するA/Dコンバータと、A/Dコンバータから出力される音声信号の特徴を抽出する特徴抽出部と、

上記音声信号から音声区間を検出する音声区間検出部と、検出された音声区間の特徴抽出されたパターンと、入力パターンとして記憶する入力パターン記憶部と、入力パターンを特徴パターンと比較して入力パターンを特定する音声認識部とを備えた音声認識装置において、

ゲイン別のホットキー信号を複数設けるとともに、ホットキー入力時、これら複数のホットキー信号を用いて検出度を各々演算し、最大検出度が得られたホットキー信号に設定されたゲインに応じて以後入力される音声に對する音声認識装置のゲインを設定する手段を備えたことを特徴とする音声認識装置。

【請求項2】 請求項(1)記載の音声認識装置において、上記マイクアンプが選択可能な複数のゲインを有し、上記最大検出度が得られたホットキー信号に設定されたゲインに対応して、マイクアンプのゲインを選択するようにしたことを特徴とするもの。

【請求項3】 請求項(1)記載の音声認識装置において、上記最大検出度が得られたホットキー信号に設定されたゲインに対応して、A/Dコンバータの参照電圧を決定するようにしたことを特徴とするもの。

【請求項4】 請求項(1)記載の音声認識装置において、ゲイン別の複数のゲインを設け、上記最大検出度が得られたホットキー信号に設定されたゲインに対応したゲインを選択するようにしたことを特徴とするもの。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、マイクrohンを通過して入力された音声を増幅する音声認識装置に関するものである。

【0002】

【発明技術】 この種音声認識装置の振幅増大率を上昇する一方で、一つの大きなネックは、音声レベルの個人差であり、音声のパワーでみると、音の大きい人と小さい人との差は数10dB程度の差がある。この音声レベルと振幅率との間には、図1に示すように、低レベルの領域では高い振幅率を得ることができず、ある程度以上のレベルでないと有効な振幅率を得ることができない。従って、低レベルから高レベルの広い範囲にわたって高い振幅率を確保することは現状では困難であり、振幅増大率としては、平均的な音声レベルで最大振幅率を得られるような設定とせざるを得ない。したがって、その場合には、低レベルの音声入力に対しては必要な振幅率が得られない。

【0003】 このため、従来においては、マイクアンプ

にAGC(自動増幅制御回路)を付加し、常に一定した音声入力を得るようにしたものが提案されている(実開59-60700号公報参照)。しかしながら、上記AGCで対応できるのは、ただか40dB程度であって、前述のように、数10dBの個人差に充分に対応することができない。

【0004】 また、発声者に適当な年度で当該発声者の音声レベルを知らせ、発声者に対し最適なレベルで発声するように促すようにしたのも提案されている(実開平01-137497号公報、特開昭63-014200号公報参照)。しかしながら、発声レベルは個人、個人特有のもので、無理に安さようとする、発声自体が不自然となり、却って振幅率を低下させるおそれがある。また、同一人の発声においても音と音とでは10dB以上のパワー差があるため、最適なレベルが設定が実際上は困難であるといった問題があった。

【0005】

【課題を解決すべき課題】 したがって、本発明の技術的課題は、発声者の音声レベルに応じて自動的に最大振幅が得られるようにすることである。

【0006】

【課題を解決するための手段】 このため、本発明は、マイクrohンと、マイクrohンの音声信号出力を増幅するマイクアンプと、マイクアンプの出力信号をデジタル信号に変換するA/Dコンバータと、A/Dコンバータから出力される音声信号の特徴を抽出する特徴抽出部と、上記音声信号から音声区間を検出する音声区間検出部と、検出された音声区間の特徴抽出されたパターンと、入力パターンとして記憶する入力パターン記憶部と、入力パターンを特徴パターンと比較して入力パターンを特定する音声認識部とを備えた音声認識装置において、ゲイン別のホットキー信号を複数設けるとともに、ホットキー入力時、これら複数のホットキー信号を用いて検出度を各々演算し、最大検出度が得られたホットキー信号に設定されたゲインに応じて以後入力される音声に對する音声認識装置のゲインを設定する手段を備えたことを特徴とする音声認識装置を提供するものである。

【0007】 即ち、本発明においては、ホットキー入力を利用して、入力されたホットキーの音声レベルに応じて、装置全体としてのゲインを最大振幅率が増えるように位置する。ここで、装置全体としてのゲインとは、音声入力を直接に増幅するマイクアンプのゲインにとどまらず、音声入力に対するゲインを間接的に制御する手段、例えば、A/Dコンバータに対して印加する参照電圧や、低レベルの領域で最大振幅率が増えるような振幅増大率を意味する。

【0008】 より具体的には、マイクアンプに対して複数のゲインを選択的に設定可能とし、最大検出度が得られたホットキー信号に設定されたゲインに応じて最適なゲインをマイクアンプに対して選択するようにして、より

く、A/Dコンバータの参照電圧を最適に設定するようにしてもよい。

[0009] また、ゲイン別の待機電圧使用量としておき、ゲイン別のホットキー待機電圧のうち、最大値が得られたもののゲインに応じて待機電圧を選択して、以後の処理を行うようにしてもよい。

[0010]

[作用・効果] 本発明によれば、ホットキー入力を利用して音声認識装置全体としてのゲインを音声者の音声レベルに応じて最適に設定することができ、発音者に自然な発音を強いることなく、常に最大値で待機電圧で音声認識を行うことができるようになる。

[0011] また、請求項2の発明によれば、ゲイン別のマイクアンプを複数用意することによって、最大値の最適なマイクアンプを選択することによって、最大値の待機電圧を確保することができる。請求項3の発明によれば、A/Dコンバータに対する参照電圧を音声者の音声レベルに応じて最適に設定することにより、最大値の待機電圧を確保することができる。

[0012] さらに、請求項4の発明によれば、ゲイン別の待機電圧を複数用意して、発音者の音声レベルに応じて待機電圧を選択することにより、最大値の待機電圧を確保することができる。

[0013]

[実施例] 以下、本発明の実施例を具体的に説明する。
(基本システム) 図2に本発明にかかる音声認識装置の基本システムを示す。マイク1から入力された音声信号は、前処理部2で前処理したうえで、マイクアンプ3を構成する自動利得制御回路(AGC)で増幅され、フイルタバンク4によって各周波数成分ごとに分離され、A/Dコンバータ5によってデジタル信号に変換され、処理部6に入力され、待機電圧が行われる。また、各周波数成分の音声信号は、音声区間検出部7にも入力され、音声区間が検出される。待機電圧された音声パターンは、検出された音声区間で区切られた状態で入力パターンとして入力パターン記憶部8に記憶される。

[0014] 記憶された入力パターンは、参照電圧部9において待機電圧10の各テンプレートと照合され、類似度算出部11において類似度が計算される。類似度算出部11は算出した類似度のうち最大値を与えるテンプレートを選択し、システム全体の類似度を行う類似度12にこれを出力する。制御部12は、入力されたテンプレートに対応する認識結果を待機電圧部13に出力し、結果出力部13はその認識結果を表示する。

[0015] 上記の音声認識装置は、所定ホットキー入力によって動作を開始するようにしており、ホットキー入力があった場合には、ゲイン別に設定された複数のホットキー待機電圧14をセレクト15によって選択し、参照電圧部9は、ホットキー待機電圧14を用いて照合を行い、最大値の待機電圧を与えるホットキー待機電圧を検出する。

[0016] 図3は、上記制御部12が実行するホットキー入力処理ルーチンを示すものであって、システム中の音量レベルを最適に設定する。

[0017] 入力パターン記憶部8に記憶された音声信号入力パターン記憶部9に入力されると、まずステップS1においてホットキー処理が終了したか否かをフラグによって判定し、ホットキー処理が終了していれば、通常の音声認識処理を実行する。フラグが立っていないと、つまりホットキー処理が終了していないときは、ステップS2において当該音声入力パターン記憶部9が読み取られ、ホットキーであるか判断された場合には、ステップS3において、ゲイン別に予め設定されたホットキー待機電圧14を用いて類似度を比較する。ゲイン別のホットキー待機電圧は、例えば、標準の音声レベルを中心として(0dB)、±6dB、±12dBの音声レベルで作成しておき、入力されたホットキーに対して最大の類似度を与えるホットキー待機電圧を検出する。ステップS4では、最大値の待機電圧を待機電圧のゲインから、発音者の音声レベルを推定し、例えば、-6dBのホットキー待機電圧が最大値の待機電圧を与えるとした場合には、ステップS5において、標準の音声レベルより6dBだけ低い音声レベルでシステム全体の待機電圧を最大となるように、システム全体としての音声レベルを設定する。この音声レベルの設定が終了すると、ステップS6でホットキー処理の終了を示すフラグをセットし、リターンする。

[0018] (システム) 音声レベル設定方式

(その1) 図4に音声レベル設定方式の一例を示す。図に示すように、この例では、5つのゲイン別ホットキー待機電圧14-1、14-2、14-3、14-4、14-5にゲイン別のマイクアンプ3-1、3-2、3-3、3-4、3-5を接続し、入力される音声レベルに応じてゲインのマイクアンプを選択するようにしている。

[0019] 図4は、標準の音声レベル0dBに対して、±12dB、±6dBの音声レベルを予測し、それぞれの音声レベルに対して最適なゲインを与えた計5個のマイクアンプ3-1、3-2、3-3、3-4、3-5を接続し、入力される音声レベルに応じてゲインのマイクアンプを選択する。この選択は、図2に示したホットキー処理において、例えば、+6dBのホットキー待機電圧14-2が最大値の待機電圧を与えると判定された場合、この待機電圧14-2が最大値の待機電圧21に記憶され、マイクアンプセレクト20はこの待機電圧21に記憶されているホットキー待機電圧の選択を取り、+6dBのホットキー待機電圧14-2である場合には、これに対応した(K+6dB)のマイクアンプ3-2を選択する。

[0020] なお、図4は図2とは対応していないが、図2で示した部分と同一の部分には同一の番号を付してそれ以上の説明を省略する。

[0021] (その2) 図5は、音声レベル設定方式の一例を示す。この場合には、A/Dコンバータ5に印加するリファレンス電圧を音声レベルに応じて選択

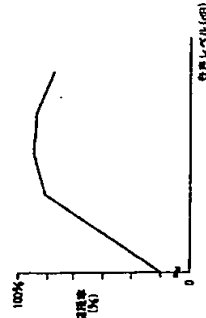
する。即ち、標準の音声レベルに対応するリファレンス電圧(0dBと表示)に対して、±12dB、±6dBのリファレンス電圧を予め選択可能に設定しておき、最も類似した待機電圧21に記憶されているホットキー待機電圧のゲインに対応したリファレンス電圧をセレクト22で選択する。なお、図5において、図4と同様、図2と同じものは同じ番号を付して説明を省略する。

[0022] (その3) 図6は音声レベル設定方式の他の例を示す。図示のように、本例では、通常の音声認識に使用する待機電圧をゲイン別に設定し、具体的には、標準の音声レベルで作成した待機電圧(0dB待機電圧)の他に、標準の音声レベルより±6dB、±12dBだけ相違する音声レベルで作成した待機電圧5つの待機電圧10-1、10-2、10-3、10-4、10-5をゲイン別にホットキー待機電圧14-1、14-2、14-3、14-4、14-5に記憶されたホットキー待機電圧が明瞭な+6dBのゲイン14-2であれば、これに対応した+6dBの待機電圧10-2をセレクト23により選択する。なお、図6において、図4と同様、図2と同じものは同じ番号を付して説明を省略する。

[図面の簡単な説明]

[図1] は入力音声レベルと類似率との関係を示すグラフ

[図1]



グラフである。

[図2] は本発明にかかる音声認識装置の概略システム図である。

[図3] は本発明において実行されるホットキー入力処理のフローチャートである。

[図4] は本発明の一実施例を示す概略説明図である。

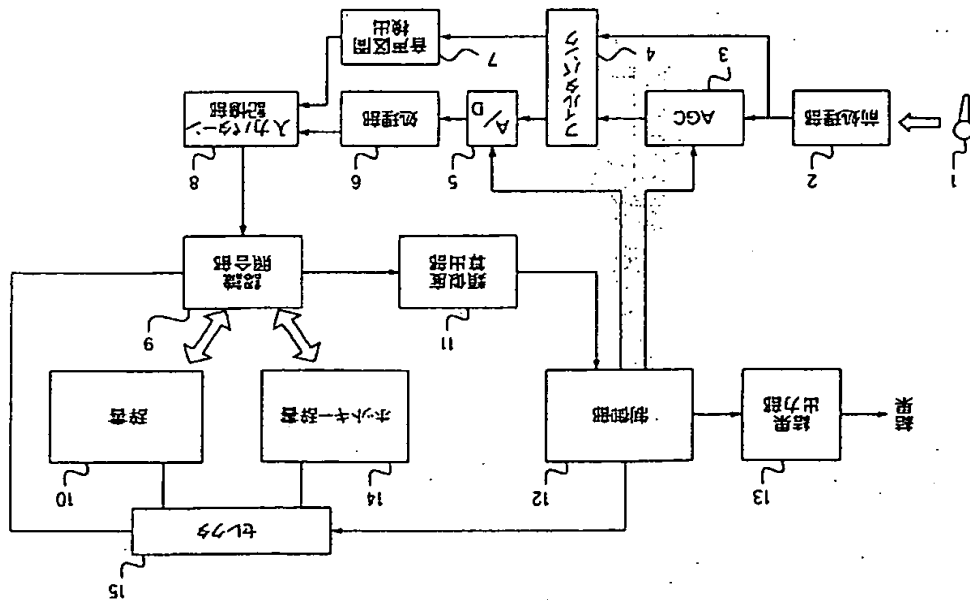
[図5] は本発明の他の実施例を示す概略説明図である。

[図6] は本発明のいまだ一つの実施例を示す概略説明図である。

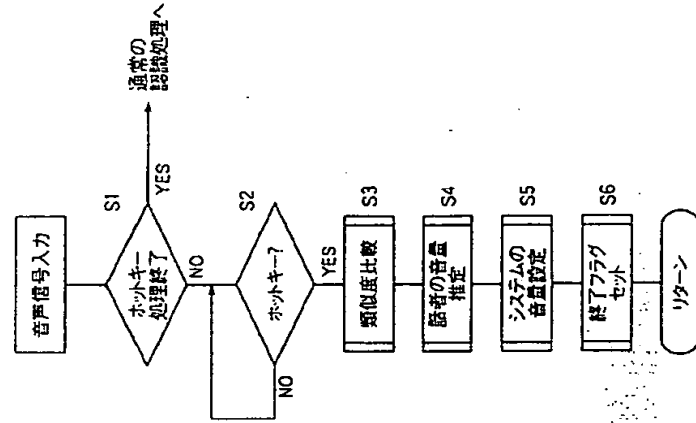
[符号の説明]

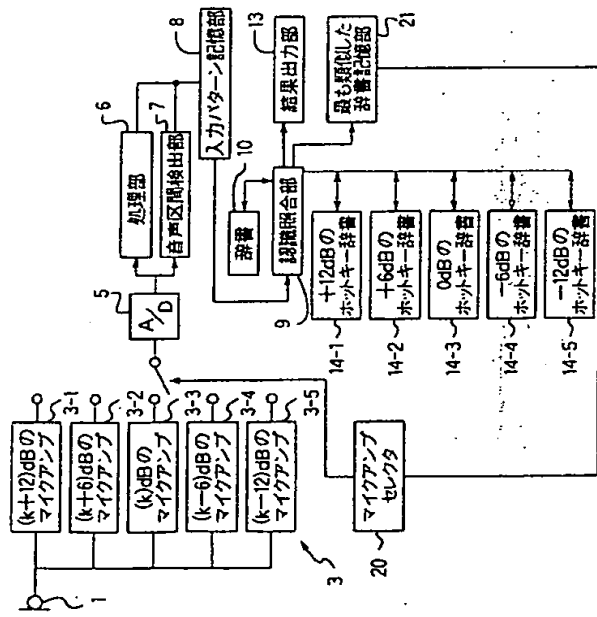
1...マイク
3-1、3-2、3-3、3-4、3-5...ゲイン別マイクアンプ
5...A/Dコンバータ
6...処理部
7...音声区間検出部
9...参照電圧部
10...待機電圧
11...類似度算出部
12...制御部
10-1、10-2、10-3、10-4、10-5...ゲイン別待機電圧
14-1、14-2、14-3、14-4、14-5...ゲイン別ホットキー待機電圧

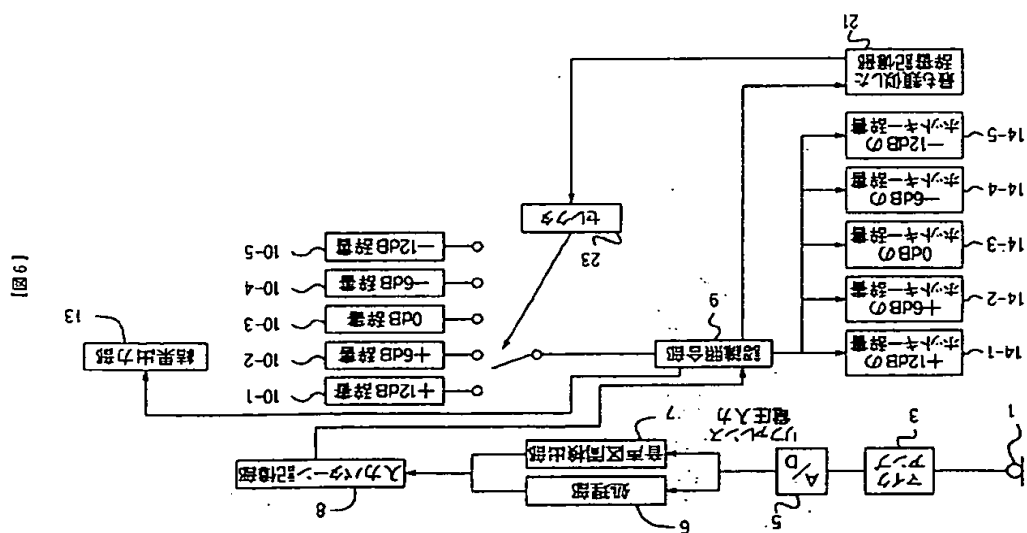
[図2]



[図3]







[98]